Stochastic Processes and their Applications 49 (1994) 141–158
North-Holland

Light traffic approximations in general stationary single-server queues

D.J. Daley *

Statistics Research Section, School of Mathematical Sciences, Australian National University, Canberra, ACT, Australia

T. Rolski **

Mathematical Institute, University of Wrocław, Poland

Received 12 August 1991
Revised 4 June 1992

This paper complements two previous studies (Daley and Rolski, 1984, 1991) by indicating the extent to which characteristics of a general stationary point process taken as the arrival process of a single-server queue influence light traffic limit theorems for the two essentially distinct schemes of dilation and thinning as routes to the limit. Properties of both the work-load and the waiting-time processes are derived, reflecting respectively the stationary time-sampling frame that may be appropriate for monitoring the system as a whole, and the customer-sampling frame (Palm distributions). Substantially different results can come from these two different views, and when compounded with the different approaches to the light traffic limits, no single light traffic scenario emerges.

AMS 1980 Subject Classifications: Primary 60K25; Secondary 60G55.

G/GI/1 queue * stationary point process with multiple points * light traffic * queue with batch arrivals

1. Introduction

This paper continues work of Daley and Rolski (1984, 1991), called (I) and (II) below. As before, our concern is with limit theorems that yield approximations to the behaviour of queueing systems in light traffic. Here we investigate the consequences of assuming that arrivals occur at the epochs of a stationary point process (not necessarily a renewal process), and study the properties of both the stationary waiting time and the stationary work load or virtual waiting-time processes in a single-server system with independent service times,

Correspondence to: Dr. T. Rolski, Mathematical Institute, Wrocław University, pl. Grunwaldzki 2/4, 50384 Wrocław, Poland.

* Work done in part while visiting the Department of Statistics, University of California, Santa Barbara in 1988.

** Work done in part while visiting the Mathematical Sciences Institute of Cornell University in 1988, supported there by U.S.A.R.O. Contract DAAG29-85-C-0018, and the School of Mathematical Sciences of the Australian National University in 1991.

SSDI 0304-4149(93)00059-H

i.e., a G/GI/1 system. The aspect of (II) concerning many-server systems is discussed elsewhere (Daley and Rolski, 1992b).

Except in Theorem 2 we allow batch arrivals. One potential application we envisage is that of highly clustered arrivals in an otherwise sparse service system. For such a system, the arrival process may be better approximated by a point process with multiple points, i.e., arrivals in batches, rather than by a simple point process. The work load in single-server queues in light traffic has also been studied by Baccelli and Brémaud (1991), Sigman (1992) and Sigman and Yamuzaki (1993), but only for simple (non-batch) arrivals.

In Theorem 3 we study light traffic for the waiting time in light traffic conditions via thinning. It is quite intuitive that in this case we do not require the input to be simple; the theorem allows us to develop light traffic results for Cox/GI/1 queues. We then obtain as special cases earlier results of Daley and Rolski (1991) on periodic queues and Burman and Smith (1986) on Markov modulated queues.

In this paper neither the batch sizes nor the time intervals between batches are assumed to be independent, other than their being independent of the service times. Since the literature on such point processes is not so readily accessible, part of Section 2 is devoted to a condensed account of the results from that area that we use, and Section 3 to the modified assumptions we then make concerning the queueing system G/GI/1 in light traffic via dilation (cf. assumptions for GI/GI/1 in (I, II)).

2. G/GI/1, its arrival process, and light traffic schemes

Since our concern is with queues under light traffic conditions, we assume without comment that the basic system from which we start is stable and exists in its stationary state.

The queueing system we investigate would be designated in Borovkov (1976, Sections 1–2) by

$$\langle \mathrm{G}, \mathrm{G}, \mathrm{G}_I, 1\rangle, \quad \text{and sometimes also} \quad \langle \mathrm{G}, 1, \mathrm{G}_I, 1\rangle; \tag{2.1}$$

we use G/GI/1 in Kendall's notation in the spirit of Stoyan (1983). We assume that customers are served in the order of their arrivals (FCFS) and that the waiting room is infinite. Specifically, the service times denoted $\{S_i\} \equiv \{S_i: i=0, \pm 1, \pm 2, \ldots\}$ constitute a family of independent identically distributed (i.i.d.) random variables (r.v.'s) with $S_i \stackrel{\mathrm{d}}{=} S$ where at the very least $ES < \infty$. The arrival process is stationary and metrically transitive, always independent of $\{S_i\}$. In the context of (2.1) it is described by the stationary sequence of nonnegative r.v.'s $\{T_i\} \equiv \{T_i: i=0, \pm 1, \pm 2, \ldots\}$ for which, in order that the system be stable, $\rho \equiv ES/ET < 1$. We allow the possibility that $\varpi \equiv \Pr\{T_i = 0\} > 0$; when $\varpi = 0$ the arrival process is a simple point process and the latter designation at (2.1) is appropriate. Baccelli and Brémaud (1987) consider only simple arrival processes, and mostly so too do Franken et al. (1982), although from the latter the relations given later in this section can be recovered using marked point processes. We choose a direct approach in the hope of making our discussion more transparent.

It is known (Loynes, 1962; cf. e.g. Borovkov, 1976, Section 3) that if

$$0 \leqslant ES < ET < \infty \tag{2.2}$$

then a stationary sequence of waiting time r.v.'s $\{W_n\}$ exists, and has the representation

$$W_n \equiv \sup_{j \geqslant 0} \left\{ \sum_{i=-j}^{-1} (S_{n+i} - T_{n+i}) \right\}, \tag{2.3}$$

and by stationarity satisfies the recurrence relation

$$W_n \stackrel{d}{=} W_{n+1} = \sup_{j \geqslant 0} \left\{ 0,\ S_n - T_n + \sum_{i=-j}^{-1} (S_{n+i} - T_{n+i}) \right\}$$
$$= (S_n - T_n + W_n)_+ \,. \tag{2.4}$$

Write $W \equiv W(\{S_i\}, \{T_i\})$ for a r.v. having the marginal distribution of each such W_n. We exclude the trivial case $ES = 0$ hereafter.

When $\{T_i\}$ is a sequence of i.i.d. r.v.'s, so that the system is GI/GI/1, it is known from Kiefer and Wolfowitz (1956) that for $\alpha > 0$,

$$\text{when (2.2) holds,} \quad EW^{\alpha} < \infty \quad \text{if and only if} \quad ES^{\alpha+1} < \infty \,. \tag{2.5}$$

In Miyazawa (1979), Wolff (1991) and Daley and Rolski (1992a) it has been indicated how properties of $\{T_i\}$ can affect the finiteness or otherwise of moments of W. It appears from these papers that for a stable G/GI/1 queueing system there is no condition for the finiteness of EW^{α} that is both analogous to (2.5) (as a necessary and sufficient condition) and has its attendant simplicity. We therefore adopt the following definition.

Definition M. The stationary ergodic sequence $\{T_i\}$ and sequence of i.i.d. r.v.'s $\{S_i\}$ (equivalently, a G/GI/1 queueing system) satisfies Condition M_{α} when

$$EW^{\alpha} \equiv E([W(\{S_i\}, \{T_i\})]^{\alpha}) < \infty \,. \tag{2.6}$$

Rather than a metrically transitive sequence $\{T_n\}$, the arrival process can be described instead as a time-stationary metrically transitive point process $N(\cdot)$ with boundedly finite first moment measure. Irrespective of $N(\cdot)$ being simple or not, there exists a one-one relation between such stationary point processes $N(\cdot)$ and stationary sequences of non-negative r.v.'s $\{T_i\}$ (Slivnyak, 1966; cf. Daley and Vere-Jones, 1988, Sections 3.4–5 and 12.3, Brandt et al., 1990, Section 7.1). It is also possible to represent the input by stationary metrically transitive bivariate sequences $\{(T_n^*, J_n^*)\}$ in which J_n^* denotes the number of customers in a batch labelled n and $T_n^* > 0$ denotes the time-interval between batches $n-1$ and n. Theorem 7.1.1 from Brandt et al. gives a one-one relationship between distributions of $\{T_n\}$ and $\{(T_n^*, J_n^*)\}$. Thus we have to distinguish three types of stationarity: (i) time stationarity, (ii) batch stationarity and (iii) customer stationarity. Except in case (iii) when $\Pr\{J_n^* = 1 \text{ (all } n)\} = 1$, and then only (ii) and (iii), none of these can co-exist with either of the others. In case (ii) we assume that batch labelled 0 arrives at $\tau_0^* = 0$ and batch n at $\tau_n^* = T_1^* + T_n^*$. In case (iii) we assume that customer zero arrives at $\tau_0 = 0$ and customer

n at τ_n, where $T_n = \tau_n - \tau_{n-1}$. Following Brandt et al. we can write the relationship between a stationary sequence $\{T_n\}$ and stationary point process $N(\cdot)$ with boundedly finite first moment measure by

$$Ef(N) = \frac{1}{ET^*} E \int_0^{T_1^*} f(S_t N^*) \, dt , \tag{2.7}$$

where $f\colon \hat{\mathcal{N}}_{\mathbb{R}} \to \mathbb{R}$ is a measurable function and S_t denotes the shift on t (see Daley and Vere-Jones, 1988, for definition of $\hat{\mathcal{N}}_{\mathbb{R}}$, the space of realizations of point processes). The relationship between $\{(T_n^*, J_n^*)\}$ and $\{T_n\}$ is simpler, namely

$$\begin{aligned} \Pr\{J_n^* = k, T_n^* \leqslant x\} &= \Pr\{J_1^* = k, T_1^* \leqslant x\} \\ &= \Pr\{x \geqslant T_k > 0 = T_{k-1} = \cdots = T_1 \mid T_0 > 0\} . \end{aligned} \tag{2.8}$$

The d.f.'s

$$R_j(x) \equiv \Pr\{T_1 + \cdots + T_j \leqslant x\} \tag{2.9}$$

determine the zero-deleted expectation function of the arrival stream, if finite, by

$$H(x) \equiv \sum_{j=1}^{\infty} R_j(x) \tag{2.10}$$

(e.g. Daley and Vere-Jones, 1988, Chapter 3). We assume throughout this paper that the point process of arrivals is such that its expectation function, defined always as the sum above of d.f.'s of partial sums of the stationary sequence $\{T_i\}$, is finite: then $H(x)/x$ converges to some finite constant as $x \to \infty$ (see e.g. Daley and Vere-Jones, 1988, Chapters 3 and 12). $1 + H$ is the analogue of the renewal function $\sum_0^\infty F^{n*}$ when $N(\cdot)$ is a stationary renewal process with generic lifetime d.f. F for which $F(0+) = 0$.

We recall the simplest relations between some of the above quantities. We have already defined $\varpi = \Pr\{T_n = 0\}$. Then

$$ET = (1 - \varpi) ET^* \quad \text{and} \quad EN(0, 1] = \frac{1}{ET} = \frac{1}{(1 - \varpi) ET^*} . \tag{2.11}$$

For $k = 1, 2, \ldots$, the batch-size distribution $\{\pi_k\} = \{\Pr\{J_n^* = k\}\}$ satisfies

$$\pi_k \equiv \Pr\{T_k > 0 = T_{k-1} = \cdots = T_1 \mid T_0 > 0\} . \tag{2.12}$$

In deducing relations like those at (2.14)–(2.15) below it can be convenient to use either or both of the identities

$$\begin{aligned} 1 &\overset{\text{a.s.}}{=} \sum_{k=1}^{\infty} I\{T_1 = \cdots = T_{k-1} = 0 < T_k\} \\ &\overset{\text{a.s.}}{=} \sum_{k=0}^{\infty} I\{T_0 = \cdots = T_{-(k-1)} = 0 < T_{-k}\} . \end{aligned} \tag{2.13}$$

Assume that $EJ_n^* = \sum_{k=1}^\infty k\pi_k < \infty$. Then

$$1-\varpi=\Pr\{T_0>0\}=\frac{1}{EJ_n^*}=1\bigg/\sum_{k=1}^{\infty}k\pi_k\,, \tag{2.14}$$

$$\varpi_j\equiv EI\{T_1=\cdots=T_{j-1}=0<T_j\}=(1-\varpi)\sum_{k=j}^{\infty}\pi_k\,. \tag{2.15}$$

Note that ϖ_j can be thought of as the probability that a randomly chosen arrival is the jth arrival in a batch.

The light traffic approximation considered in (I) and part of (II) is based on dilating the time scale of the arrival process by γ. In terms of the point process $N(\cdot)$ and the stationary sequence of intervals $\{T_i\}$, such a *γ-dilation* amounts to their being replaced by

$$N^{(\gamma)}(\cdot)\equiv N(\gamma^{-1}\cdot)\quad\text{and}\quad\{\gamma T_i\} \tag{2.16}$$

respectively. A second family of light traffic approximations considered in (II) and most of the literature cited in the references envisages independent thinning of the arrival process, which is equivalent to replacing $\{T_i\}$ by the sequence $\{T_i^{(\pi)}\}$ defined by

$$T_i^{(\pi)}=\sum_{j=n_i}^{n_{i+1}-1}T_j\,, \tag{2.17}$$

where $n_0=0$, $n_{i+1}=n_i+\nu_i$ $(i\geqslant 0)$, $n_i=n_{i+1}-\nu_i$ $(i<0)$, and $\{\nu_i\colon i=0,\pm 1,\pm 2,\ldots\}$ is a set of i.i.d. r.v.'s geometrically distributed on $\{1,2,\ldots\}$ with $\Pr\{\nu=k\}=\pi(1-\pi)^{k-1}$. Such *$\pi$-thinning* (or, *$\pi$-deletion*) yields a point process $N^{(\pi)}(\cdot)$ whose evaluation $N^{(\pi)}(A)$ on any bounded Borel set A is related to its pre-deletion value $N(A)$ by $N^{(\pi)}(A)\overset{\mathrm{d}}{=}B(N(A),\pi)$ where $B(n,\pi)$ denotes a binomial r.v. with mean $n\pi$. Later at (4.1) we write $n_i=n_i^{(\pi)}$.

It seems too trite to note that, for a Poisson process, the operations of dilation and thinning are stochastically equivalent. Yet we should, because it follows from the properties of ergodic point processes under thinning and rescaling (e.g. Daley and Vere-Jones, 1988, Section 9.3) that it is only for Poisson arrival processes that we should expect the effects of the deterministic operation of γ-dilation and the random operation of π-thinning to coincide when $\pi^{-1}=\gamma$ for every service distribution with $E(S^2)<\infty$. So, as noted in (II), how we choose the sequence of processes involved in a light traffic approximation is a non-trivial matter. In practical terms this means that *the choice of which set of light traffic results should be applied in particular circumstances is a non-trivial decision.*

3. Waiting times in G/GI/1 in light traffic via dilation

For the light traffic approximation given by dilation as at (2.16), consider

$$W(\gamma)\equiv W(\{S_i\},\{\gamma T_i\}) \tag{3.1}$$

for large γ, as in (I), (II) and Whitt (1988). Observe from (2.16) and (2.4) that $W(\gamma)$ is a.s. nonincreasing in γ. Use the identity at (2.13) in the form

$$1 \overset{\text{a.s.}}{=} \sum_{k=1}^{\infty} I\{\gamma T_{-1} = \cdots = \gamma T_{-k+1} = 0 < \gamma T_{-k}\}\,, \tag{3.2}$$

in conjunction with the representation for $W(\gamma)$ that comes from (2.4) and (3.2). Then for $\gamma \to \infty$,

$$W(\gamma) \xrightarrow{\text{a.s.}} W(\infty) \equiv \sum_{k=1}^{\infty} (S_{-1} + \cdots + S_{-k+1}) \times I\{T_{-1} = \cdots = T_{-k+1} = 0 < T_{-k}\}\,. \tag{3.3}$$

The indicator r.v.'s $I\{\cdot\}$ here are independent of $\{S_i\}$, and from (2.15) they have expectations

$$\varpi_j = EI\{T_1 = \cdots = T_{j-1} = 0 < T_j\}\,.$$

Our first light traffic result is informative when $\varpi \equiv \Pr\{T=0\} > 0$.

Theorem 1. *In a stationary metrically transitive* G/GI/1 *queueing system in which* $S \overset{\text{a.s.}}{>} 0$ *the stationary waiting time r.v.'s* $W(\gamma)$ *of the family defined by* γ*-dilation satisfy*

$$\lim_{\gamma\to\infty} \Pr\{W(\gamma) > 0\} = \varpi\,. \tag{3.4}$$

Without the restriction that $\Pr\{S>0\} = 1$,

$$\lim_{\gamma\to\infty} \Pr\{W(\gamma) > 0\} = 1 - \sum_{k=1}^{\infty} \varpi_k (\Pr\{S=0\})^{k-1} = \sum_{k=1}^{\infty} [1 - (\Pr\{S=0\})^{k-1}]\,\varpi_k\,. \tag{3.5}$$

When condition M_1 *is satisfied,* $\lim_{\gamma\to\infty} EW(\gamma)$ *exists and is finite, being given by*

$$\lim_{\gamma\to\infty} EW(\gamma) = \sum_{k=2}^{\infty} (k-1) ES\varpi_k = (1-\varpi) ES \sum_{j=2}^{\infty} \tfrac{1}{2} j(j-1)\,\pi_j\,. \tag{3.6}$$

Proof. The proofs follow from (3.3) and the monotone convergence theorem.

For (3.4), under the assumption that $S \overset{\text{a.s.}}{>} 0$ we have

$$\Pr\{W(\infty) > 0\} = \Pr\left\{ \bigcup_{k=2}^{\infty} \{T_{-1} = \cdots = T_{-k+1} = 0 < T_{-k}\} \right\} = 1 - \Pr\{T_{-1} > 0\} = \varpi\,.$$

For (3.5),

$$\Pr\{W(\infty)>0\}=\Pr\left\{\bigcup_{k=1}^{\infty}\bigcup_{j=1}^{k}\{S_{-k+1}>0\}\cap\{T_{-1}=\;=T_{-k+1}=0<T_{-k}\}\right\}$$
$$=\sum_{k=1}^{\infty}[1-(\Pr\{S=0\})^{k-1}]\varpi_k . \qquad \square$$

In the context of an arrival process consisting of 'rarely occurring' clusters, as for example with occasional overflows from another system, identify Pr{cluster contains k arrivals} with π_k at (2.12). Then Theorem 1 gives an approximation to the mean waiting time per arrival, though this gives little idea as to the range of waiting times to be observed, for which the ratio of (3.6) by (3.4), i.e., $E(W|W>0)$, is a better indicator. If it holds that the service times are rather larger than the inter-arrival times T'_i within a cluster, then a better approximation than (3.6) is

$$EW\approx(1-\varpi)(ES-ET')\sum_{j=2}^{\infty}\tfrac{1}{2}j(j-1)\pi_j . \tag{3.7}$$

This can be formally justified by the following proposition. In it we consider an R/GI/1 queue introduced by Wolff (1991), namely, a single-server queue with the stationary inter-arrival times $\{T_n(\gamma)\}$ induced by a regenerative process with a typical cycle of the following form. Suppose that $T'_1, T'_2, \ldots$ is a sequence of non-negative i.i.d. r.v.'s independent of the positive integer-valued random variable J. Then J is the cycle length and a typical cycle is $T'_1, \ldots, T'_{J-1}, \gamma T'_J$. In (3.7) $\pi_j\approx\Pr\{J=j\}$.

Proposition 1. *Let an* R/GI/1 *queue have stationary regenerative input* $\{T_n(\gamma)\}$, *for which* J *and* $\{T'_i\colon 1\leqslant i\leqslant J-1\}$ *are independent,* $ET'_i\equiv ET'$ *and*

$$T'_i\overset{\text{a.s.}}{\leqslant}S_i . \tag{3.8}$$

Then, when condition M_1 *holds,*

$$\lim_{\gamma\to\infty}EW(\gamma)=(ES-ET')\frac{\sum_{j=2}^{\infty}(j-1)j\Pr\{J=j\}}{2EJ} . \tag{3.9}$$

Proof. Suppose that $\{T_i^\circ(\gamma)\}$ is regenerative with generic cycle $T'_1, \ldots, T'_{J-1}, \gamma T'_J$. There exists a sequence $\{W_n^\circ(\gamma)\}$ satisfying

$$W_{n+1}^\circ(\gamma)=(W_n^\circ(\gamma)+T_n^\circ(\gamma)-S_n)_+ .$$

We have with probability 1,

$$W_0^\circ(\gamma)\to 0 \quad\text{and}\quad W_J^\circ(\gamma)\to 0 \qquad (\gamma\to\infty) . \tag{3.10}$$

By Proposition 2.3 of Rolski (1981),

$$EW(\gamma)=\frac{E\sum_{j=0}^{J-1}W_j^\circ}{EJ} .$$

Bearing in mind (3.8) and (3.10), the limit of $EW(\gamma)$ for $\gamma \to \infty$ equals

$$\frac{E \sum_{j=1}^{J-1} \sum_{i=1}^{j} (S_i - T_i')}{EJ},$$

and now standard calculations yield (3.9). □

Theorem 1 gives a valid statement, albeit uninformative, about the limit behaviour of $W(\gamma)$ when $\varpi = 0$. Theorem 1 of (II) gives rate of convergence information when

$$u^{-\alpha} \Pr\{T_1 \leqslant u\} \to c_A \quad (u \downarrow 0) \tag{3.11}$$

for some positive finite α and c_A, namely, for GI/GI/1 queues,

$$\text{if } E(S^{\alpha+1}) < \infty \quad \text{then } \lim_{\gamma \to \infty} \gamma^\alpha \Pr\{W(\gamma) > 0\} = E(S^\alpha) c_A, \tag{3.12a}$$

$$\text{if } E(S^{\alpha+2}) < \infty \quad \text{then } \lim_{\gamma \to \infty} \gamma^\alpha EW(\gamma) = \frac{E(S^{\alpha+1}) c_A}{\alpha + 1}. \tag{3.12b}$$

We now give sufficient conditions under which similar results hold for G/GI/1 when the arrival process is a simple point process.

Denote by $(\Omega, \mathscr{E}, \mathscr{P})$ a probability space supporting the stationary metrically transitive sequence $\{(S_i, T_i): i = 0, \pm 1, \ldots\}$, and let $\mathscr{F}_0$ denote a sub-σ-field containing $\mathscr{G}_0 = \sigma\{T_i: i = -1, -2, \ldots\}$. Then there exists a regular conditional probability $P(\cdot \mid \mathscr{F}_0)(\cdot)$ such that

$$\mathscr{P}(\{T_0 \leqslant u\} \cap B) = \int_B P(\{T_0 \leqslant u\} \mid \mathscr{F}_0)(\omega) \mathscr{P}(\mathrm{d}\omega) \quad (\text{all } B \text{ in } \mathscr{F}_0), \tag{3.13}$$

(see e.g. Breiman, 1968, Section 4.3). Assume that, for some $\alpha > 0$ and $\delta > 0$, $P(\cdot \mid \mathscr{F}_0)(\cdot)$ is such that

$$P(\{T_0 \leqslant u\} \mid \mathscr{F}_0)(\omega) \leqslant K(\omega) u^\alpha \quad (0 < u \leqslant \delta) \tag{3.14a}$$

for some random variable $K(\cdot)$ and that the limit

$$C_A(\omega) \equiv \lim_{u \to 0} u^{-\alpha} P(\{T_0 \leqslant u\} \mid \mathscr{F}_0)(\omega) \tag{3.14b}$$

exists a.s. In this paper we assume that K is bounded. Note that the random variables C_A and S_0 are independent. Then because of (3.14a),

$$\begin{aligned} c_A \equiv \lim_{u \to 0} u^{-\alpha} \Pr\{T_0 \leqslant u\} &= \int_\Omega \lim_{u \to 0} u^{-\alpha} P(\{T_0 \leqslant u\} \mid \mathscr{F}_0)(\omega) \mathscr{P}(\mathrm{d}\omega) \\ &= \int_\Omega C_A(\omega) \mathscr{P}(\mathrm{d}\omega). \end{aligned} \tag{3.15}$$

We now adapt to the G/GI/1 context the arguments used in establishing (3.12) for the case of a renewal arrival process.

Theorem 2. *In a stationary metrically transitive* G/GI/1 *queueing system which satisfies Condition* M_α *and whose arrival process is a simple point process satisfying* (3.14) *for some finite positive* α, *the stationary waiting time r.v.'s* $W(\gamma)$ *defined by* γ*-dilation satisfy*

$$\lim_{\gamma\to\infty} \gamma^\alpha \Pr\{W(\gamma)>x\} = c_{\mathrm{A}} E([(S-x)_+]^\alpha) . \tag{3.16}$$

When Condition $\mathrm{M}_{\alpha+1}$ *is satisfied moreover, the right-hand side below is finite and*

$$\lim_{\gamma\to\infty} \gamma^\alpha EW(\gamma) = \frac{E(S^{\alpha+1})c_{\mathrm{A}}}{\alpha+1} . \tag{3.17}$$

Proof. Write $A(x\,|\,\omega) = P(\{T_0 \leqslant x\}\,|\,\mathscr{F}_0)(\omega)$, $A_-(x\,|\,\omega) = P(\{T_0 < x\}\,|\,\mathscr{F}_0)(\omega)$. Using the notation at (3.1) with (2.4), we have from (3.13) that for any $x \geqslant 0$,

$$\begin{aligned}\mathscr{P}(\{W_1(\gamma)>x\}) &= \mathscr{P}(\{W_0(\gamma)+S_0-\gamma T_0 > x\}) \\ &= \int_\Omega A_-(\gamma^{-1}[W_0(\gamma)+S_0-x]\,|\,\omega)\mathscr{P}(\mathrm{d}\omega) .\end{aligned}$$

As in (II) for GI/GI/1, the a.s. monotonicity in γ of $(W_0(\gamma)\,|\,\mathscr{F}_0)$ and the existence of the limit $C_{\mathrm{A}}(\cdot)$ implies that for $\gamma \geqslant \gamma'$ for sufficiently large γ',

$$\gamma^\alpha A_-(\gamma^{-1}[W_0(\gamma)+S_0-x]\,|\,\omega) \leqslant [(W_0(\gamma')+S_0-x)_+]^\alpha K \quad \text{a.s.} \tag{3.18}$$

for some finite random variable K. When Condition M_α is satisfied, $E\{[(W_0(\gamma')+S_0-x)_+]^\alpha\} < \infty$ and we can then use the dominated convergence theorem with the convergence to 0 of $W(\gamma)$ to complete the proof. To prove (3.17) write

$$\begin{aligned}E(W_1(\gamma)) &= E((W_0(\gamma)+S_0-\gamma T_0)_+) \\ &= \int_\Omega \left(\frac{S_0+W_0}{\gamma}\right)^{\alpha+1} \frac{1}{((S_0+W_0)/\gamma)^{\alpha+1}} \\ &\quad \times \int_0^{(S_0+W_0)/\gamma} A(x\,|\,\omega)\ \mathrm{d}x\ \mathscr{P}(\mathrm{d}\omega) .\end{aligned}$$

Similarly as in (II) we can prove that as $\gamma\to\infty$,

$$\begin{aligned}&\int_\Omega \left[\left(\frac{S_0+W_0}{\gamma}\right)^{\alpha+1} \frac{1}{((S_0+W_0)/\gamma)^{\alpha+1}} \int_0^{(S_0+W_0)/\gamma} A(x\,|\,\omega)\ \mathrm{d}x\right] \mathscr{P}(\mathrm{d}\omega) \\ &\quad \to C_{\mathrm{A}}(\omega)E(S^{\alpha+1}) ,\end{aligned}$$

and the proof is completed by the dominated convergence theorem. $\square$

Example 1. *Alternating renewal process.* Suppose that the inter-arrival times are those of an alternating renewal process in which the respective lifetime d.f.'s $A_j(\cdot)$ $(j=1,2)$, say, are given by $A_j(t)=t^j$ $(0\leqslant t\leqslant 1)$. Then the condition at (3.14a) is satisfied with $\delta=1$, $\alpha=1$, and $C_A(\omega)=1$ or 0 depending on the lifetime being of the first or second type. For such a process the constant $c_A=\frac{1}{2}$. In this example we may take $\mathscr{F}_0=\mathscr{G}_0$ because by the strong law of large numbers $\mathscr{G}_0$ determines whether A_1 or A_2 is the distribution of T_0.

The inter-arrival process in Example 1 is a special case of a Markov renewal process. This example is readily extended to Markov renewal processes or Markov modulated arrival processes, but then instead of $\mathscr{F}_0=\mathscr{G}_0$ we must take a sub-σ-field $\mathscr{F}_0$ which includes not only $\mathscr{G}_0$ but also a history for the generating Markov mechanism.

4. Waiting times in G/GI/1 in light traffic via thinning

For the family of light traffic approximations obtained by π-thinning, the generic stationary waiting time r.v. has the representation (cf. (2.4) and (2.17))

$$W_0^{(\pi)}=\sup_{j\geqslant 1}\left\{0,\left(\sum_{i=-j}^{-1}S_i-\sum_{i=n_{-j}^{(\pi)}}^{-1}T_i\right)\right\}$$

$$\stackrel{\mathrm{d}}{=}W_1^{(\pi)}=(S_0-T_0^{(\pi)}+W_0^{(\pi)})_+ \tag{4.1}$$

where $S\stackrel{\mathrm{d}}{=}S_0$ is independent of $\{S_i\}$, $T_0^{(\pi)}$ and $W_0^{(\pi)}$, and $T_0^{(\pi)}=\sum_{j=1}^{\nu_0}T_{j-1}$ for $\nu_0\stackrel{\mathrm{d}}{=}\nu\equiv\nu(\pi)$ geometrically distributed on 1, 2, ... with mean $1/\pi$. Moreover, these $\{\nu_i(\pi)\}$ defined at (2.17) are independent, independent of $\{T_i\}$, and increase in distribution to ∞ as $\pi\downarrow 0$. Thus it is possible to find a probability space supporting r.v.'s $\bar{n}_{-j}^{(\pi)}$ that are equivalent in distribution to $n_{-j}^{(\pi)}$ and $\downarrow -\infty$ monotonically a.s. as $\pi\to 0$. Take the product of this space with the space supporting the defining sequences $\{T_i\}$, $\{S_i\}$, so that for every j, $\sum_{i=\bar{n}_{-j}^{(\pi)}}^{-1}T_i\to\infty$ a.s. $(\pi\to 0)$, and monotonically so. Then, observing that $W_0^{(\pi)}$ is a decreasing function of the partial sums of the T_i and that with respect to $\pi\to 0$ the S_i are fixed, it follows that

$$\Pr\{W^{(\pi)}=0\}\to 1 \quad \text{as } \pi\to 0\,. \tag{4.2}$$

Theorem 3. *In a stationary metrically transitive* G/GI/1 *queueing system satisfying Condition* M_1, *the stationary waiting time r.v.'s* $W^{(\pi)}$ *of the light traffic approximations derived by independent* π*-thinning of the arrival process satisfy*

$$\lim_{\pi\to 0}\pi^{-1}\Pr\{W^{(\pi)}>x\}=EH_-(S-x)\,, \tag{4.3}$$

where $H_-(u)=\sum_{i=1}^{\infty}\Pr\{T_1+\cdots+T_i<u\}$ *is the left-continuous version of the (zero-deleted) expectation function of the Palm distribution for the arrival point process. When Condition* M_2 *is satisfied,*

$$\lim_{\pi \to 0} \pi^{-1} EW^{(\pi)} = E\left(\int_0^S H_-(u) \, du \right). \tag{4.4}$$

Remark. The result at (4.3) is slightly more general than the analogue at Theorem 2 of (II), given there only for $x = 0$. The argument is similar whether $x =$ or > 0.

Proof of Theorem 3. Use the identity $\sum_{r=1}^{\infty} I\{\nu_0 = r\} \overset{\text{a.s.}}{=} 1$ together with the representation at (4.1) to write

$$W^{(\pi)} \overset{\mathrm{d}}{=} W_1^{(\pi)} = \sum_{r=1}^{\infty} I\{\nu = r\}(S_0 - T_0^{(\pi)} + W_0^{(\pi)})_+ , \tag{4.5}$$

so for $x \geqslant 0$,

$$\begin{aligned} \Pr\{W^{(\pi)} > x\} &= \sum_{r=1}^{\infty} \Pr\{\nu_0 = r\} \Pr\{T_0 + \cdots + T_{r-1} < S_0 + W_0^{(\pi)} - x\} \\ &= \pi \sum_{r=1}^{\infty} (1-\pi)^{r-1} \Pr\{T_0 + \cdots + T_{r-1} < S_0 + W_0^{(\pi)} - x\} \\ &\geqslant \pi \sum_{r=1}^{\infty} (1-\pi)^{r-1} \Pr\{T_0 + \cdots + T_{r-1} < S_0 - x\} . \end{aligned}$$

By the finiteness assumption on $H_-(\cdot)$ below (2.10) and its asymptotic linearity, the right-hand side below is finite when $ES < \infty$, so we have the relation

$$\begin{aligned} \liminf_{\pi \to 0} \pi^{-1} \Pr\{W^{(\pi)} > x\} &\geqslant \sum_{r=1}^{\infty} \Pr\{T_0 + \cdots + T_{r-1} < S_0 - x\} \\ &= EH_-(S_0 - x) . \end{aligned} \tag{4.6}$$

For an inequality in the reverse direction, we let π' denote some fixed but small positive value for π. Then for $\pi \leqslant \pi'$,

$$\begin{aligned} \pi^{-1} \Pr\{W^{(\pi)} > x\} &\leqslant \sum_{r=1}^{\infty} (1-\pi)^{r-1} \Pr\{T_0 + \cdots + T_{r-1} < S_0 + W_0^{(\pi')} - x\} \\ &\leqslant \sum_{r=1}^{\infty} \Pr\{T_0 + \cdots + T_{r-1} < S_0 + W_0^{(\pi')} - x\} \qquad (4.7) \\ &\leqslant \sum_{r=1}^{\infty} \Pr\{T_{-r} + \cdots + T_{-1} < S_{-r} + W_{-r}\} \\ &= E \sum_{r=1}^{\infty} I\{T_{-r} + \cdots + T_{-1} < S_{-r} + W_{-r}\} \equiv EL_0 , \qquad (4.8) \end{aligned}$$

where L_0 denotes the number of customers in service or waiting for service at the arrival epoch of the 0th customer. The waiting time of this customer, if non-zero, can be represented as the sum of the residual service time $\hat{S}$ of the customer in service and the service times of the other $L_0 - 1$ customers waiting in the queue. Hence

$$W_0 = \hat{S} + \sum_{i=1}^{L_0-1} S_i \overset{\mathrm{d}}{\geqslant} \sum_{i=1}^{L_0-1} S_i \tag{4.9}$$

where the r.v.'s $\{S_i : i = 1, 2, \dots\}$ are independent of L_0 and $\overset{\mathrm{d}}{\geqslant}$ denotes the stochastic ordering relation. Condition M_1 yields

$$\infty > EW_0 \geqslant E(L_0 - 1)ES ,$$

implying that the sum at (4.7) is finite. Consequently, letting $\pi' \to 0$, by (4.2), the continuity property of probability measure and the monotone convergence theorem,

$$\begin{aligned}
&\sum_{r=1}^{\infty} \Pr\{T_0 + \cdots + T_{r-1} < S_0 + W_0^{(\pi')} - x\} \\
&\quad = \sum_{r=1}^{\infty} \Pr\{T_0 + \cdots + T_{r-1} < S_0 - x;\ W_0^{(\pi')} = 0\} \\
&\qquad + \sum_{r=1}^{\infty} \Pr\{T_0 + \cdots + T_{r-1} < S_0 + W_0^{(\pi')} - x;\ W_0^{(\pi')} > 0\} \\
&\quad \to EH_-(S_0 - x) .
\end{aligned}$$

To prove (4.4), write

$$\begin{aligned}
\pi^{-1} EW^{(\pi)} &= \sum_{r=1}^{\infty} (1-\pi)^{r-1} E(S_0 + W_0^{(\pi)} - T_0 - \cdots - T_{r-1})_+ \qquad (4.10) \\
&\geqslant \sum_{r=1}^{\infty} (1-\pi)^{r-1} E(S_0 - T_0 - \cdots - T_{r-1})_+ \\
&\to \sum_{r=1}^{\infty} E(S_0 - T_0 - \cdots - T_{r-1})_+ \\
&= \sum_{r=1}^{\infty} E \int_0^S \Pr\{T_0 + \cdots + T_{r-1} \leqslant u\} \, \mathrm{d}u \\
&\geqslant E \int_0^S H_-(u) \, \mathrm{d}u ,
\end{aligned}$$

establishing (4.4) as a lower bound. Also, much as in the argument around (4.7), we have

$$\pi^{-1}EW^{(\pi)} \leqslant \sum_{r=1}^{\infty} E(W_0^{(\pi')} + S_0 - T_0 - \cdots - T_{r-1})_+ \tag{4.11}$$

$$\leqslant \sum_{r=1}^{\infty} E(W_{-r} + S_{-r} - T_{-r} - \cdots - T_{-1})_+$$

$$= E\int_0^{\infty} \sum_{r=1}^{\infty} I\{T_{-r} + \cdots + T_{-1} \leqslant W_{-r} + S_{-r} - t\}\, dt$$

$$= E\int_0^{\infty} L_0(t)\, dt$$

where $L_0(t)$ denotes the number of those of the L_0 customers arriving before the arrival epoch of the 0th customer that are still present at a time t later. Using the notation at (4.9) this integral can be rewritten as

$$E\left(L_0\hat{S} + \sum_{i=1}^{L_0-1} (L_0 - i)S_i\right),$$

and this is finite provided that both $E(L_0\hat{S})$ and $E(L_0^2)$ are finite. But (4.9) and Condition M_2 yield

$$\infty > E(W_0^2) = E(\hat{S}^2) + 2ESE[\hat{S}(L_0-1)] + E(L_0-1)E(S^2)$$
$$+ E[(L_0-1)(L_0-2)](ES)^2,$$

so the required finiteness conditions are satisfied. As before, letting $\pi' \downarrow 0$ and using monotonicity in (4.11) yields (4.4), proving the theorem. □

We emphasize that Theorem 3, unlike Theorem 2, holds irrespective of the arrival process being simple or not. In the former case the expectation function has a simpler representation, though in the latter event the more complicated expression simplifies if the variance function for $N(\cdot)$ is known (cf. Daley and Vere-Jones, 1988, Section 3.5).

In the next theorem we consider a single-server queue with arrivals according to a Cox process $N_{\lambda*}(\cdot)$ having random intensity function $\{\lambda^*(t): -\infty < t < \infty\}$ which is a stationary ergodic non-negative random process with finite intensity

$$0 < \bar{\lambda} = E[\lambda^*(0)] < \infty,$$

and finite cross-product function

$$a(t) = \frac{E[\lambda^*(t)\lambda^*(0)]}{\bar{\lambda}} < \infty. \tag{4.12}$$

Notice that thinning such a Cox process with retention probability π results in a Cox process with random intensity function $\{\pi\lambda^*(t): -\infty < t < \infty\}$. Assume further that λ^* has trajectories in $D(-\infty, \infty)$, i.e. they are right-continuous and have left-hand limits. We use Cox/GI/1 to denote such a single-server queue with i.i.d. service times which are independent of the input. For such queues, conditions for M_α have been given in terms of the strong mixing coefficient function in Daley and Rolski (1992a).

Theorem 4. *In a* Cox/GI/1 *queue, provided Condition* M_1 *(respectively,* M_2*) is satisfied,*

$$\lim_{\pi\to 0} \pi^{-1} \Pr\{W^{(\pi)}>0\} = \int_0^\infty \Pr\{S>x\}a(x)\, dx \tag{4.13a}$$

and

$$\lim_{\pi\to 0} \pi^{-1} E[W^{(\pi)}] = \int_0^\infty E(S-x)_+\, a(x)\, dx\,. \tag{4.13b}$$

Proof. Define a process $\{\lambda^0(t): -\infty<t<\infty\}$ with trajectories in $D(-\infty, \infty)$ by

$$\Pr\{\lambda^0 \in \cdot\} = E\left(\int_0^1 \frac{I\{\tau^s\lambda^* \in \cdot\}\lambda^*(s)\, ds}{\bar{\lambda}}\right), \tag{4.14}$$

where τ^s denotes the shift operator. It follows from Grandell (1976, Sections 2.4.1 and 2.4.2) that the synchronous point process corresponding to N_{λ^*} is the point process $\delta_0 + N_{\lambda^0}$, where N_{λ^0} is a Cox process with intensity function λ^0.

To verify (4.13), we apply Theorem 3, for which we find $H_-(\cdot)$ as follows:

$$\begin{aligned} H_-(x) = E(N_{\lambda^0}(0, x)) &= \int_0^x E[\lambda^0(t)]\, dt \\ &= \int_0^x dt \int_0^1 \frac{E[\lambda^*(s+t)\lambda^*(s)]\, ds}{\bar{\lambda}} \quad \text{(by (4.14))} \\ &= \frac{1}{\bar{\lambda}} \int_0^x a(t)\, dt\,. \quad \square \end{aligned}$$

Corollary 1 (cf. (II), Section 5). *In a* G/GI/1 *queue with periodic Poisson arrivals with periodic intensity function* $\lambda(\cdot)$ *of period* σ *and* $ES^2<\infty$ *(respectively* $ES^3<\infty$*),* (4.13a) *and* (4.13b) *hold respectively with* $\lambda^*(t) = \lambda(\theta^*+t)$, *where* θ^* *is a r.v. uniformly distributed on* $(0, \sigma)$.

Proof. Condition M_1 (respectively, M_2) is fulfilled provided $ES^2<\infty$ (respectively, $ES^3<\infty$); see Daley and Rolski (1992a) and references therein. $\square$

Corollary 2 (cf. Burman and Smith, 1986, Theorem 2.1). *In a* G/GI/1 *queue with Markov-modulated arrival process, with intensity of the form $f(X(t))$, where $X(t)$ is a finite state irreducible Markov process and $ES^2 < \infty$ (respectively, $ES^3 < \infty$),* (4.13a) *and* (4.13b) *hold respectively with $\lambda^*(t) = f(X(t))$, where $X(t)$ is a stationary version of an irreducible continuous time Markov chain on finite state space.*

Proof. Condition M_1 (respectively, M_2) is fulfilled provided $ES^2 < \infty$ (respectively, $ES^3 < \infty$); see Daley and Rolski (1992a). □

5. Virtual waiting time or work load

In this section we allow arrivals in batches. For general stationary $N(\cdot)$, whether simple or not, and adopting the convention that $N(0, a] = -N(a, 0]$ for $a < 0$, we have the representation

$$V(t) \equiv \sup_{s \leqslant t} \left\{ \sum_{i=N(0,s]+1}^{N(0,t]} S_i - (t-s) \right\} \tag{5.1}$$

of the work load at time t (see e.g. Borovkov, 1976, Section 6). This is analogous to the functional at (2.3) for the actual waiting time, and for a first-come first-served queue discipline $V(t)$ is the virtual waiting time. Equivalently, we can define a new system with a simple input and having the same work-load process, namely, the times of arrivals in this new system of (single) customers are just the arrival times of batches in the old system, and every arriving customer at the new system has as its service time the total service time of all customers in the batch in the old system. However, unless the original input is either simple or a renewal process, the service times in the new system are no longer independent nor independent of the inter-arrival times.

In both these systems, the stationary work-load processes in these two systems are in fact identical, and can be represented as at (5.1) with $S = S_n^*$ and $N(\cdot)$ suitably defined. We use S_n^* and W_n^* to denote the service time and the (stationary) waiting time of the arrival labelled n in the new system, noting that in general they are not independent. The stationary distributions for W_n and W_n^* are related as at (7.3.2) of Brandt et al. (1990) from which it follows that $W_n \overset{\mathrm{d}}{\geqslant} W_n^*$, and hence that for given $\alpha > 0$, $E(W_n^\alpha) < \infty$ implies $E[(W_n^*)^\alpha] < \infty$.

Following e.g. Borovkov (1976) or Franken et al. (1982),

$$\Pr\{V(t) > 0\} = ES/ET . \tag{5.2}$$

Consequently, for the processes $V^{(\gamma)}(\cdot)$ and $V^{(\pi)}(\cdot)$ defined by γ-dilation and π-thinning respectively,

$$\Pr\{V^{(\gamma)}(t) > 0\} = ES/ET^{(\gamma)} = \gamma^{-1}(ES/ET) , \tag{5.3}$$

$$\Pr\{V^{(\pi)}(t) > 0\} = ES/ET^{(\pi)} = \pi(ES/ET) , \tag{5.4}$$

as observed by Whitt (1988) for (5.3). Beyond noting the equality of these two probabilities when $\pi = \gamma^{-1}$, nothing else need be said about these quantities. Recall that $m = EN(0, 1] = [(1-\varpi)ET^*]^{-1}$. The following theorem illustrates the impact of allowing inputs with multiple points on the behaviour of characteristics in light traffic conditions. Equation (5.5) below is a special case of Sigman's (1992) Theorem 2.1 which gives the limit of $\gamma \Pr\{V^{(\gamma)}(t) > x\}$ in the case of a simple input process and service time sequence $\{S_i\}$ that need not be i.i.d.

Theorem 5. *In a stationary metrically transitive* G/GI/1 *queueing system the stationary work-load process* $V^{(\gamma)}(t)$ *of the light traffic approximation defined by* γ*-dilation satisfies*

$$\gamma \Pr\{V^{(\gamma)}(t) > 0\} = mES. \tag{5.5}$$

When Condition M_2 *is satisfied,*

$$\lim_{\gamma\to\infty} \gamma E(V^{(\gamma)}(t)) = m(1-\varpi) \sum_{i=1}^{\infty} \tfrac{1}{2}\pi_i [iE(S^2) + i(i-1)(ES)^2]. \tag{5.6}$$

Proof. Equation (5.5) follows directly from (5.3). To prove (5.6) we have (Brumelle, 1971, Theorem 8; or Rolski, 1981, Corollary 5.2)

$$EV(t) = \frac{1}{2ET^*} (E(W_n^* + S_n^*)^2 - E(W_n^*)^2), \tag{5.7}$$

where

$$S_n^* = \sum_{j=R_n+1}^{R_{n+1}} S_j \quad (n = \dots, -1, 0, 1, \dots)$$

with

$$R_i = \begin{cases} J_1^* + \cdots + J_i^* & (i = 0, 1, \dots), \\ -(J_0^* + J_{-1}^* + \cdots + J_i^*) & (i = -1, -2, \dots). \end{cases}$$

Under γ-dilation, $W_n^* \downarrow 0$ a.s., so $\gamma EV^{(\gamma)}(t) \to \frac{1}{2}m(1-\varpi)E[(S^*)^2]$, and

$$E(S^*)^2 = \sum_{j=1}^{\infty} E\left(\sum_{i=1}^{j} S_i\right)^2 \pi_j = \sum_{i=1}^{\infty} \tfrac{1}{2}\pi_i [iE(S^2) + i(i-1)(ES)^2],$$

which completes the proof. $\square$

Observe that, while we have used Condition M_2 in proving the result at (5.6), it is only the weaker Condition M_1 that is implicitly involved in the result. An analogous observation is possible for several other light traffic results: it is particularly apt to give it here because Condition M_1 is neither implicit in (5.5) nor is it used in the proof above.

For light traffic conditions defined by π-thinning, we have a different result for the first moment.

Theorem 6. *In a stationary metrically transitive* G/GI/1 *queueing system the stationary work-load process* $V^{(\pi)}(t)$ *of the light traffic approximation defined by* π*-thinning satisfies*

$$\pi^{-1}\Pr\{V^{(\pi)}(t)>0\}=mES\,. \tag{5.8}$$

When Condition M_2 *holds,*

$$\lim_{\pi\to 0}\pi^{-1}E(V^{(\pi)}(t))=\tfrac{1}{2}mE(S^2)\,. \tag{5.9}$$

Proof. Equation (5.8) follows from (5.4). The proof of (5.9) is similar to that of (5.6). Thus under π-thinning, for $\pi\downarrow 0$, W_n^* is decreasing stochastically in distribution to 0, $\pi ET_n^*\to ET$ and $ES_n^*\to ES$. To prove (5.9) we apply (5.7) similarly as in the proof of (5.6). □

One way of visualizing the difference between the limits of Theorems 5 and 6 is via the sample paths of $V(t)$, a typical realization of which is a saw-tooth function. The saw-teeth in its π-thinned path $V^{(\pi)}(t)$ consist, with a few rare exceptions, of shifted versions of realizations of the function $(S-t)_+I_{\mathbb{R}_+}(t)$, and the probability of two saw-teeth being located close together is $O(\pi)$. On the other hand under γ-dilation, with a few rare exceptions each saw-tooth is a shifted version of a realization of

$$\sum_{i=1}^{N_0(\{0\})}(S_{-i}-t)_+I_{\mathbb{R}_+}(t)\,.$$

These saw-teeth are fundamentally different in nature when $\Pr\{N_0(\{0\})>1\}>0$, i.e. the case of non-trivial batch arrivals. Even in the case of simple $N_0(\cdot)$ they are juxtaposed differently, though not enough to make any difference to the formulae at (5.6) and (5.9).

The contrast between (5.6) and (5.9), while not as marked as between any of (3.6), (3.12b) and (4.4), still serves as a reminder that *for light traffic limits, batch effects in the arrival process persist under* γ*-dilation but are destroyed under* π*-thinning.*

Acknowledgement

We thank Ward Whitt and Volker Schmidt for their interest and comments. We thank the referee for a meticulous reading of our original manuscript: several improvements have resulted.

References

F. Baccelli and P. Brémaud, Palm Probabilities and Stationary Queues. Lecture Notes in Statist. No. 41 (Springer, Berlin, 1987).

F. Baccelli and P. Brémaud, Virtual customers in sensitivity and light traffic analysis via Campbell's formula for point processes, manuscript (1991).

A.A. Borovkov, Stochastic Processes in Queueing Theory (Springer, New York, 1976).
A. Brandt, P. Franken and B. Lisek, Stationary Stochastic Models (Akademie, Berlin, and Wiley, Chichester, 1990).
L. Breiman, Probability (Addison-Wesley, Reading, MA, 1968).
D.Y. Burman and D.R. Smith, An asymptotic analysis of a queueing system with Markov-modulated arrivals, Oper. Res. 34 (1986) 105–119.
D.J. Daley and T. Rolski, A light traffic approximation for a single-server queue, Math. Oper. Res. 9 (1984) 624–628.
D.J. Daley and T. Rolski, Light traffic approximations in queues, Math. Oper. Res. 16 (1991) 57–71.
D.J. Daley and T. Rolski, Finiteness of waiting-time moments in general stationary single-server queues, Ann. Appl. Probab. 2 (1992a) 987–1008.
D.J. Daley and T. Rolski, Light traffic approximations in many-server queues, Adv. Appl. Probab. 24 (1992b), 202–218.
D.J. Daley and D. Vere-Jones, An Introduction to the Theory of Point Processes (Springer, New York, 1988).
P. Franken, D. König, U. Arndt and V. Schmidt, Queues and Point Processes (Akademie, Berlin, and Wiley, Chichester, 1982).
J. Grandell, Doubly Stochastic Poisson Processes. Lecture Notes in Math. No. 529 (Springer, Berlin, 1976).
J. Kiefer and J. Wolfowitz, On the characteristics of the general queueing process with application to random walks, Ann. Math. Statist. 27 (1956) 147–161.
R.M. Loynes, The stability of a queue with non-independent inter-arrival and service times, Proc. Cambridge Philos. Soc. 58 (1962) 497–520.
M. Miyazawa, A formal approach to queueing processes in the steady state and their applications, J. Appl. Probab. 16 (1979) 332–346.
T. Rolski, Stationary Random Processes Associated with Point Processes. Lecture Notes in Statist. No. 5 (Springer, New York, 1981).
K. Sigman, Light traffic for work-load in queues, Queueing Syst. 11 (1992) 429–442.
K. Sigman and G. Yamuzaki, Heavy and light traffic in fluid models with burst arrivals, Ann. Inst. Statist. Math. 45 (1993) 1–7.
I.M. Slivnyak, Stationary streams of homogeneous events, Vestnik Har'kov. Gos. Univ. Ser. Mech.-Math. 32 (1966) 73–116.
D. Stoyan, Comparison Methods for Queues and Other Stochastic Processes (Wiley, Chichester, 1983) [English ed., D.J. Daley, ed.]
W. Whitt, An interpolation approximation for the mean workload in a GI/G/1 queue, Oper. Res. 37 (1988) 936–952.
R.W. Wolff, On finite delay-moment conditions in queues, Oper. Res. 40 (1991) 771–775.